PROPERTY OF
NCETON UNIVERSITY LIBRARY
RECEIVED MAY 22 1939

The
PSYCHOLOGICAL
RECORD . . . .

MAY, 1939
Vol. III No. 7

AN EMPIRICAL STUDY OF VERBAL, NUMERICAL, AND SPATIAL FACTORS IN MENTAL ORGANIZATION

ISIDOR CHEIN
College of the City of New York

THE PRINCIPIA PRESS, INC.
BLOOMINGTON, INDIANA

Price of this number, 60 cents

EDITOR — J. R. KANTOR, *Indiana University*
BUSINESS EDITOR — C. M. LOUTTIT, *Indiana University*

DEPARTMENTAL EDITORS:

ABNORMAL — EDMUND S. CONKLIN, *Indiana University*
CHILD — HELEN KOCH, *University of Chicago*
CLINICAL — B. M. CASTNER, *Yale University*
COMPARATIVE — E. A. CULLER, *University of Rochester*
EDUCATIONAL — J. G. PEATMAN, *City College of New York*
EXPERIMENTAL — B. F. SKINNER, *University of Minnesota*
MEDICAL — NORMAN CAMERON, M. D., *Cornell Medical College*
PHYSIOLOGICAL — C. F. SCOFIELD, *University of Buffalo*
PSYCHOMETRICS — J. P. GUILFORD, *University of Nebraska*
SOCIAL — NORMAN C. MEIER, *University of Iowa*
ASSISTANT EDITOR — J. W. CARTER, JR., *Indiana University*

The Principia Press, Inc., has undertaken the publication of this co-operative journal to afford authors the opportunity of immediate publication at the least possible expense. The present low cost of publication, and possible future reductions, depend entirely upon the number of subscriptions. The subscription price has been made low to induce individuals to subscribe. Under the Articles of Incorporation of the Principia Press no profit can be made on any of its publications. Therefore an increase in the number of subscribers will be reflected in reduced costs to authors and in increase in the number of pages published annually.

EDITORIAL BOARD. The above named board of associate editors have agreed to serve in an active capacity for a period of three years, and all manuscripts will be submitted to them.

MANUSCRIPTS may be sent to the appropriate associate editor or to Dr. J. R. Kantor. Longer papers (30 or more typewritten pages) in the above mentioned fields are especially desired.

COSTS. The cost of publication will be $2.00 per page of approximately 470 words. The charge for line cut illustrations will be $2.00 regardless of size. Special charges will have to be made for complex tabular matter and for half-tone or colored illustrations.

REPRINTS. One hundred copies of each paper will be supplied gratis. Additional copies may be purchased in any quantity at a rate which will be set when the author is informed of the cost of publication.

ROYALTIES. Fifty per cent of the net income from the sale of individual copies, or from copies sold as part of back volumes, will be credited as royalties to the author's account. Royalties cannot be paid on income from subscriptions current in the year of publication.

SUBSCRIPTIONS. The subscription price is $4.00 per year for an annual volume of approximately 500 pages. Individual papers will be sold at an advertised price, which will depend upon the size of the article. Foreign subscriptions will be $4.50.

CORRESPONDENCE concerning editorial matters should be addressed to Dr. J. R. Kantor, Indiana University. Business correspondence should be addressed to Dr. C. M. Louttit, Principia Press, Inc., Bloomington, Ind.

CHAPTER I*†

Problem and Procedure

The Problem

Methods of factor analysis represent a highly useful technique for studying the organization of abilities. Concretely this has meant the reduction of test performances in a wide variety of situations to a relatively small number of experimentally determined and operationally defined reference abilities such as verbal ability, numerical ability and the like. The value of such simplification for scientific psychology is, of course, immeasurable. Then, again, in such studies as those of Flanagan (5) and Guilford (8) factor analysis has proven its value in the field of personality despite the fact that for various reasons this method is probably "not especially well adapted to the study of personality" (6, p. 295). Finally, factor analysis has proved useful as a technique in test construction (1, 2, 3, 9, 16).

A large proportion of the research carried out along these lines was done before the development of the more general methods of factor analysis were developed. Garrett (6) has, however, reworked five of the earlier studies, making use of Thurstone's method with axis rotation. He found that the reality of the traits (factors) reported on the basis of the earlier studies was substantiated in "a more clear-cut fashion" and also that the newer methods of factor analysis were able "to give, in addition, new information with respect to minor factors not hitherto identified."

The only attempt to validate directly and empirically the factors themselves as psychological variables was made by Schneck (10) who studied the relationship of factor scores to scholastic per-

*This is one of a series of studies under the general direction of Professor H. E. Garrett. The writer wishes to express his appreciation and gratitude to Professor Garrett for the latter's guidance and advice. He also wishes to express his gratitude to Professor G. M. Smith, who made available to the present writer the test results for the subjects of the present study, and to Dean Morton Gottschall and Recorder J. K. Ackley of the College of the City of New York, who graciously permitted the writer access to the record files of the college.

† Recommended for publication by Dr. J. R. Kantor, April 24, 1939.

formance. He found the regression equations of the tests on the central factors in a battery of verbal and in a battery of numerical tests. From these he estimated numerical and verbal factor scores. He then found for each student the average grade in 'literary' (English, French, German, etc.) courses and the average grade in science courses. Dividing his students into A, B, and C groups for the literary and science courses, he compared these groups for the average factor scores. The results indicated a trend towards correspondence between the verbal factor and ability in literary courses, the relationship to ability in the science courses being less marked. The numerical factor showed a very slight trend toward correspondence with ability in the science courses and none at all in the literary courses. Unfortunately, he does not present critical ratios nor adequate information for computing them.

Schneck's study, however, made use of the then best available, Spearman's, technique. Although Spearman's method may still have utility for certain problems, it has largely been supplanted by the more general and, as we have indicated, more satisfactory methods of factor analysis which have since been developed. These newer methods have never been submitted to such study although there is now an increasing practice of making use of factor scores. The closest approach to an attempt at direct empirical validation of factors derived by the newer methods is given by Thurstone (15) who presents the factor profiles of a few individuals and compares them with the vocational interests of these subjects.

Research on factor analysis may now proceed along two main lines: (1) improvement of methods of factor analysis by increasing the stability and reliability of the factors, improving the psychological identification of the factors and the like, and extension of methods of factor analysis to personality problems, to the better analysis of intelligence, etc.—and (2) investigation of the relation of experimentally determined factors with other psychological variables with which one is more directly concerned in daily life.

It is the purpose of the present study to determine the relation of verbal, numerical, and spatial factors to scholastic achievement. Such study should serve the double purpose of further validating the factors and of throwing more light on the determinants of successful scholastic achievement.

Procedure

The basic data of this study, which consisted of the scores made by 186 students on fourteen tests, were very kindly provided by Dr. G. M. Smith. For the present study, the scholastic records of 178 of these subjects were obtained from the Recorder's Office at the College of the City of New York. A thorough report of the test material is given by Smith in his monograph (11) so that only a brief summary is necessary here. The test intercorrelations and reliability coefficients are given in Table A.

1. The subjects

The 186 subjects constituted a relatively homogeneous group with respect to age, sex, socio-economic status, and formal education. All were male students at the College of the City of New York, 80% of them being either in the upper sophomore or junior classes. The average age was 19 years and five months with a standard deviation of 16 months.

2. The tests and testing procedure

The fourteen tests used are fully described by Smith (loc. cit.) There were five verbal tests (sentence completion, verbal analogies, grammatical analogies, verbal generalizations, and grammatical generalizations), four numerical tests (number series completion, numerical analogies, arithmetic reasoning, and numerical generalizations) four spatial tests (Minnesota Form Board, spatial analogies, spatial generalizations, and a modification of Kelley's spatial test), and an anagrams test. The reliability coefficients for these tests varied from .689, for grammatical generalizations, to .962 for the modified Kelley spatial test (Table A). The total testing time represented 5 hours, 36 minutes, spread over a period of about 8 weeks. Smith attests to the fact that "There was every indication that the students, with very few exceptions, accepted the tests as a challenge to their intellectual powers and put forth their best efforts" (p. 19). For the convenience of the reader, sample items from each of the tests are given in Appendix A.

TABLE A

TEST INTERCORRELATIONS (RAW) AND RELIABILITY COEFFICIENTS [FROM SMITH (11)*]

Variable:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Sentence Completion	.789	.428	.176	.533	.105	.449	.346	.431	.354	.331	.370	.410	.257	.304
2. Number Series Completion		.889	.254	.487	.442	.308	.533	.577	.313	.368	.437	.474	.464	.505
3. Minnesota Form Board			.811	.263	.237	.115	.476	.305	.266	.306	.251	.129	.481	.663
4. Verbal Analogies				.852	.335	.531	.505	.562	.272	.510	.406	.520	.389	.413
5. Numerical Analogies					.884	.315	.489	.428	.354	.251	.475	.326	.339	.392
6. Grammatical Analogies						.797	.376	.421	.368	.331	.414	.590	.312	.210
7. Spatial Analogies							.753	.602	.321	.347	.499	.348	.493	.585
8. Arithmetic Reasoning								.865	.407	.373	.489	.457	.432	.462
9. Anagrams									.891	.267	.398	.344	.348	.395
10. Verbal Generalizations										.781	.410	.377	.472	.443
11. Numerical Generalizations											.830	.463	.490	.397
12. Grammatical Generalizations												.689	.361	.282
13. Spatial Generalizations													.745	.487
14. Modified Kelley Spatial														.962

*Coefficients are given by Smith to four decimal places.

3. *Factor scores*

For the purposes of the present study, the battery of fourteen tests was subjected to Thurstone's method of factor analysis (13) and the centroid axes obtained were submitted to orthogonal transformation. Multiple correlations of the tests with the three of the four analyzed factors which seemed to correspond, respectively, to verbal, numerical and spatial factors and multiple regression equations of the tests with these factors were determined. The raw test scores for each test were transmuted into equivalent scores in 'standard' normal distributions with means of 50 and standard deviations of 14. From the regression equations and the 'standardized' test scores of the subjects, verbal, numerical and spatial factor scores were estimated for the subjects.

4. *Grade scores and criteria*

Grade scores were determined by assigning values of 1, 2, 3, 4, and 5 to grades A, B, C, D, and F, respectively, multiplying each of these values by the number of credits in the course, summing the products and dividing by the total number of credits involved.

Eight categories of grade scores were used:

1. Total grades.
2. Social science—including economics, education, government, sociology, history, philosophy, psychology.
3. Science—including biology, chemistry, engineering, geology, physics.
4. Mathematics.
5. Science and mathematics, combined.
6. Foreign language—including Romance languages and Classical languages.
7. English and literature (referred to simply as English).
8. Foreign language and English, combined.

Criterion groups. For each of these grade categories, the thirty (approximately one-sixth of the total number of subjects) best and the thirty poorest students were selected. These groups constituted the criterion groups in terms of which the discriminative value of the factor scores was determined.

CHAPTER II

Results

1. Interrelationships among criterion groups

It should be apparent that if the same thirty individuals constituted each of the upper criterion groups, and the same thirty individuals constituted each of the lower criterion groups, only one comparison between the criterion groups would be justified. For whatever would be the difference between the members of any particular pair of criterion groups, this difference would also necessarily obtain between the members of each of the other pairs. Conversely, if there were no overlap at all between the criterion groups, each of the eight comparisons would be completely independent of any of the others. This latter situation would, of course, arise only if there were a zero intercorrelation between each of the scholastic categories. With intermediate degrees of overlap, consistent discriminations for the different pairs of criterion groups are separate and independent tests of the relation of the factor scores to scholastic achievement only to the extent of the degree of overlap within the upper and within the lower criterion groups, the higher the degree of overlap, the lower being the degree of independence between the comparisons.

Table I presents the percentage of overlap, or community, as compared with the highest possible community between criterion groups. That is, since each criterion group consists of 30 individuals, 100 percent community between two groups would mean that the same thirty individuals composed both groups; a percent community of zero would mean that none of the individuals in one group is to be found in the other; a percent community of 50 means that the two groups have 15 individuals in common.

It will be seen that, as is to be expected, individuals in upper groups tend to remain in upper groups (i.e., individuals who are superior in some subjects tend to be superior in others) and vice versa, although there are a considerable number of individuals who fall into both upper and lower criterion groups. It will also be seen that, on the whole, there is greater overlap in the upper criterion groups than in the lower ones (i.e., the tendency of the indi-

TABLE I

PERCENT COMMUNITY OF TOTAL POSSIBLE COMMUNITY BETWEEN CRITERION GROUPS

Criterion groups	+2	+3	+4	+5	+6	+7	+8	—2	—3	—4	—5	—6	—7	—8
+1	50	67	60	70	57	53	60	3	10	3	3	0	7	0
+2		37	33	33	40	50	47		17	20	10	13	0	3
+3			40	73	37	37	37			3	0	3	3	7
+4				70	40	33	50				0	7	13	3
+5					37	40	43					7	10	7
+6						33	73						13	0
+7							50							3
—1	0	0	0	0	3	3	3	53	40	53	47	37	43	43
—2		3	7	7	17	10	13		33	37	33	23	37	30
—3			0	0	17	13	13			53	80	37	20	30
—4				0	3	10	7				67	37	27	30
—5					7	10	7					33	27	33
—6						17	0						20	73
—7							7							40

Criterion groups are designated by their numbers: the positive numbers represent the superior students in the category represented by the number; the negative numbers represent the inferior students; 1 is total grades, 2 is social science, 3 is science, 4 mathematics, 5 science and mathematics, combined, 6 foreign language, 7 English, 8 foreign language and English, combined.

viduals who are inferior in some subjects to be inferior in others is not as great as is the tendency for individuals who are superior in some subjects also to be superior in others). This may perhaps be explained by the weeding out of generally poor students by entrance requirements and by the scholastic requirements of the college.

The greatest percent community (80%) exists between the negative science and the negative combined science and mathematics groups, the two having 24 individuals in common. The corresponding upper groups had 73 percent community, or 22 individuals in common. This means that 46 of the 60 individuals involved in the comparison of the upper and lower science groups are also included

in the 60 individuals involved in the comparison of the upper and lower combined mathematics and science groups. To this extent, then, the two comparisons are duplicating one another.

The next highest percentage of community (73%) is to be found between the upper foreign language and the upper combined foreign language and English groups and also between the corresponding lower groups. This means that 44 of the 60 individuals involved in the comparison of the superior and inferior foreign language students are also included in the 60 individuals involved in the comparison of the superior and inferior students as measured by foreign language and English considered together.

The third highest percentage of community (70%) is found between the upper groups selected on the basis of grades in science and mathematics and on the basis of total grades. The same percentage of community is found between superior students selected on the basis of performance in both science and mathematics and on the basis of performance in mathematics alone. There is 67 percent community between the lower mathematics and the lower combined science and mathematics groups. In other words, 41 of the individuals involved in the comparison of the upper and lower mathematics groups are also involved in the comparison of the superior and inferior combined science and mathematics groups. There is only 47 percent community between the lower total grades group and the lower science and mathematics group. This means that 35 of the 60 individuals involved in the comparison of the upper and lower total grades groups are also included in the 60 individuals involved in the comparison of the upper and lower combined science and mathematics groups. There is also 3 percent community between the upper total grades group and the lower science and mathematics group (i.e., there is one individual who is in both of these groups). This individual would tend to lower the discrimination between the better and poorer students selected on the basis of total grades if he tends to raise the discrimination between better and poorer students selected on the basis of mathematics and science courses alone; and vice versa.

Of special interest to us is the overlap between positive and negative groups. As we have indicated in the preceding paragraph,

an individual who is in an upper group for one scholastic category and in a lower group for another scholastic category would, if his factor scores tend to support discriminations for one of these categories, tend to lower the discriminations for the other category. Naturally, the larger the number of such individuals, the greater the handicap for satisfactory discrimination of all of the scholastic criteria by the factor scores. If, for any pair of scholastic categories, this handicap is fairly large, but both categories are nonetheless satisfactorily discriminated by the factor scores, the argument for the validity of the factor scores is thereby strengthened. Incidentally, it may have already been noticed that where the community between like signed groups is fairly large, there tends to be very low, or zero, community between the corresponding unlike signed groups.

For the social science and foreign language categories there are 9 such 'paradoxical' individuals; for the foreign language and the English categories there are also 9; for the social science and mathematics, 8; mathematics and English, 7; social science and science, 6; science and foreign language, 6.

2. *Factor loadings*

Table II presents the centroid factor analysis (13) of the fourteen tests. The average residual correlation after the extraction of the first four factors is only .029 with a standard deviation of .024. This low value and the low percentages of variance contributed by the third and fourth factors justify us in not attempting to extract any further factors. The average discrepancy between the experimental and guessed communalities is only .044 with a standard deviation of .033. Discrepancies of this order are sufficiently small not to require a repetition of the factor analysis with a closer approximation of the guessed to the experimental communalities (6, p. 292). The common factor variance in this battery of tests is only about 55 percent, the remaining variance being attributable chiefly to specific factors and errors of measurement.

The centroid axes represent arbitrary reference points and, as has been emphasized by Thurstone (14) and Guilford (7), should not (as they stand) be subjected to psychological interpretation.

Thus, if we attempt to submit the arbitrary centroid factors to psychological interpretation, we would have to postulate for our first factor a psychological variable which contributes most to performance on an arithmetic reasoning test and, in order from the next greatest to the lowest contribution, to a spatial analogies test, a verbal analogies test, the modified Kelley spatial test, a number series completion test, a grammatical generalizations test, a numerical generalizations test, etc. It is apparent that even if we postulated a general factor running through our fourteen variables, this general factor would have no systematic relationship to either the material or structural components of the tests. For our second factor, we would have to postulate a psychological variable which, while contributing favorably to performance on test of grammatical analogies, sentence completion, grammatical generalizations and verbal analogies, would have no bearing on a test of verbal generalizations and would be a *handicap* to performance on the Minnesota Form Board, the modified Kelley spatial, spatial generalizations and spatial analogies. Such an as-

TABLE II

CENTROID FACTOR LOADINGS

PROJECTIONS OF THE TESTS ON ARBITRARY CENTROID AXES AS DETERMINED BY THURSTONE'S METHOD

Test	I	II	III	IV	h^2
1. Sentence completion	.565	+.375	+.254	+.070	.529
2. No. series completion	.693	—.026	+.119	—.172	.525
3. Minnesota Form Board	.515	—.424	+.170	+.255	.539
4. Verbal analogies	.706	+.224	+.155	+.157	.598
5. Numerical analogies	.559	—.176	—.195	—.344	.499
6. Grammatical analogies	.599	+.440	—.195	+.099	.600
7. Spatial analogies	.733	—.197	+.173	—.165	.633
8. Arithmetic reasoning	.736	+.068	+.134	—.240	.622
9. Anagrams	.541	+.068	—.138	—.062	.320
10. Verbal generalizations	.595	—.019	—.051	+.277	.433
11. Numerical generalizations	.674	—.023	—.267	—.116	.539
12. Grammatical generalizations	.637	+.335	—.179	+.046	.552
13. Spatial generalizations	.654	—.270	—.147	+.153	.545
14. Mod. Kelley spatial	.697	—.413	+.223	+.120	.720
$\Sigma\frac{K^2}{N}$	.410	.072	.032	.034	.548

sumption of a *handicapping* ability would be an extreme one in the light of all we know of the correlation of abilities. Our necessary assumption for the third factor would be even more puzzling, even if we were prepared to assume handicapping abilities. We would here have to assume an ability which is favorable for sentence completion, the modified Kelley spatial test, spatial analogies, the Minnesota Form Board, verbal analogies, arithmetic reasoning and number series completion, but a handicap to numerical generalizations, numerical analogies, grammatical analogies, grammatical generalizations, spatial generalizations and anagrams. Our fourth factor would involve the assumption of a psychological variable favorable to verbal generalizations, the Minnesota Form Board, verbal analogies and spatial generalizations and a handicap to tests of numerical analogies, arithmetic reasoning, number series completion and spatial analogies.

All that the centroid analysis accomplishes is to segregate the tests in a multi-dimensional conceptual space. The problem now is to find psychologically meaningful reference axes. This is accomplished by the rotation of axes. The criteria of psychological meaningfulness, it is apparent from our previous discussion, are to find reference axes on which (a) the test projections have systematic relationships to the material or structural components of the tests and (b) negative projections are a minimum. It should be noted that in the case of personality, as opposed to ability, variables this last criterion is not so important.

With these criteria in mind, we therefore submitted the factor analysis to the following series of orthogonal transformations, all of the rotations being in a counter-clockwise direction: factors I and II were rotated 37 degrees around the others; III and IV were similarly rotated 14 degrees; I′ and IV′, 32 degrees; II′ and IV″, 83 degrees; I″ and III′, 24 degrees; IV‴ and I‴, 4 degrees. The resulting factor pattern is presented in Table III. Further orthogonal transformations resulted in a picture which was less psychologically meaningful (i.e., the relation of the factor loadings to the material or structural components of the test were less systematic and/or the negative loadings were larger or more numerous). We also attempted a further oblique transformation

TABLE III

FACTOR PATTERN AFTER ORTHOGONAL TRANSFORMATION

	Test	I_R	II_R	III_R	IV_R	h^2
1.	Sentence completion	.633	.355	—.006	.041	.528
2.	No. series completion	.325	.501	.076	.403	.525
3.	Minnesota Form Board	.267	—.012	—.109	.675	.539
4.	Verbal analogies	.657	.300	.094	.257	.597
5.	Numerical analogies	—.015	.478	.305	.422	.500
6.	Grammatical analogies	.578	.271	.438	.016	.600
7.	Spatial analogies	.277	.488	.001	.564	.633
8.	Arithmetic reasoning	.365	.601	.094	.344	.622
9.	Anagrams	.276	.304	.293	.256	.320
10.	Verbal generalizations	.475	.053	.205	.404	.434
11.	Numerical generalizations	.239	.368	.433	.400	.540
12.	Grammatical generalizations	.520	.314	.412	.115	.552
13.	Spatial generalizations	.282	.109	.260	.622	.546
14.	Mod. Kelley Spatial	.317	.201	—.104	.754	.720
	$\Sigma \frac{K^2}{N}$	.169	.125	.063	.190	.547

by Thurstone's method of averages (13, pp. 175-7), but this also resulted in a less meaningful pattern. We do not deem it necessary to cite the numerous transformations attempted before the series here reported was decided upon.

Table III, then, represents the most satisfactory pattern, from a psychological point of view, we could derive from our data.

Thurstone (13, pp. 155-6) has described three conditions which are probably necessary and sufficient for that unique simple structure which is desirable from both the statistical and psychological points of view. When there are four trait vectors, these conditions are: (1) each row of the table must contain at least one zero, (2) each column must contain four or more zeros, and (3) for every pair of columns, there must be at least four tests which have zero weights in one column, but not in the others. An examination of Table III shows that these conditions are not met. Thurstone, however, while seeking to achieve a unique mathematical solution, seeks, on the psychological side, to meet the criteria we have set for ourselves (cf. 12, pp. 247-9). Mathematically, simple structure establishes distinct clusters among the tests with an elimination of

overlap. An examination of Table IV which gives the percentage of variance in each test which is attributable to each of the factors (derived simply by squaring each of the entries of Table III) shows that, while we do not meet Thurstone's criteria, we do have reasonably distinct clusters. It should be recalled, incidentally, that this battery of tests was not devised by Smith with the intention of measuring primary abilities.

The tests with the highest saturations in the first factor are, in order of size of factor weights, variables 4, 1, 6, 12, and 10. These are the five verbal tests. The test with the next highest factor weight for this factor is variable 8, the arithmetic reasoning test, which probably also involves verbal ability to some extent. Each of the verbal tests, moreover, has its highest saturation in the first factor and relatively low weighting in the other factors. We may, therefore, fairly safely describe the first factor as a verbal factor.

TABLE IV

PERCENTAGES OF VARIANCE IN EACH TEST ATTRIBUTABLE TO EACH FACTOR

(Derived by squaring the entries in Table III)

	Test	I_R^2	II_R^2	III_R^2	IV_R^2
1.	Sentence completion	.401	.126	.000	.002
2.	No. series completion	.106	.251	.006	.162
3.	Minnesota Form Board	.071	.000	.012	.456
4.	Verbal analogies	.432	.090	.009	.066
5.	Numerical analogies	.000	.228	.093	.178
6.	Grammatical analogies	.334	.073	.192	.000
7.	Spatial analogies	.077	.238	.000	.318
8.	Arithmetic reasoning	.133	.361	.009	.118
9.	Anagrams	.076	.092	.086	.066
10.	Verbal generalizations	.226	.003	.042	.163
11.	Numerical generalizations	.057	.135	.187	.160
12.	Grammatical generaliz.	.270	.099	.170	.013
13.	Spatial generalizations	.080	.012	.068	.387
14.	Mod. Kelley Spatial	.100	.040	.011	.569

The tests having the highest saturations in the second factor are, in order of size of factor weights, variables 8, 2, 7, 5, and 11. With the exception of variable 7, the spatial analogies test, these are all numerical tests. Variables 8, 2, and 5 also have their highest saturations in the second factor. Variable 7, however, has its highest factor weight in factor four. In view of these considerations, we feel justified in describing factor two as a numerical factor. It is possible that if we had extracted a fifth factor, the further axis rotations made possible thereby would have lowered the factor loading of variable 7 in factor two and increased the relative loading of variable 11 for this factor. For reasons already described, however (p. 76f), we did not feel justified in extracting a fifth factor. Or, as seems more likely, the ambiguous position of variable 7 may be due to an accident of the sampling, since the factor weights are based on the correlations and the correlations are, of course, subject to accidents of sampling.

Factor four, it seems apparent, may be safely described as a spatial factor. The four spatial tests (variables 14, 3, 13, 7) have the four highest saturations in this factor. Variable 5, numerical analogies, which comes next in rank for size of the fourth factor weight, owes 14 percent less of its variance to this factor than does variable 7, which is the lowest of the four spatial tests in fourth factor weights. Each of the four spatial tests, moreover, has its highest saturation in the fourth factor.

Factor three is probably a structural factor. The variance attributable to this factor is only about 6% as compared to 17%, 13%, and 19% for factors I, II, and IV, respectively. This is consistent with Smith's finding (11) that the material bonds among the tests are much stronger than the structural bonds. The four variables having the highest weights are, in order, 6, 11, 12, and 5, two analogies and two generalizations tests. The remaining two generalizations tests, variables 10 and 13, are included within the next three variables in the ranking for size of loading in this factor. This factor may perhaps be interpreted as a relational factor. From the present analysis, it would seem that there is a relatively close relationship between the kind of mental operation involved in generalizations and in perceiving analogies. With the

exception of the anagrams test, which Smith classifies as a 'construction' test, the remaining construction tests, variables 2, 1, 14, and 3, occupy ranks 10, 12, 13, and 14 with respect to weighting for this factor. This is inconsistent with Smith's finding, by means of mean-tetrad analysis, that there is a greater cleavage between the analogies and generalizations tests than between either and the constructions group. It may, however, be recalled that Smith found that the members of the constructions group had less in common with one another than did the members of either the generalizations or analogies groups. It may be that the extraction of further factors would clarify the present inconsistency, although that is aside from our present interest.

The multiple correlations of the tests used in estimating the factor scores and the multiple regression equations of these tests on the factors are given in Table V.[1] Because of the ambiguous position of variable 7, we felt that despite the fact that it would serve to raise the multiple correlation of the tests with the numerical factor, it would be psychologically more meaningful to disregard it in our estimation of the numerical factor scores.

3. *Discrimination of criterion groups by factors*

The mean factor scores, the differences, standard deviations of the differences, and critical ratios of the differences between

[1] The multiple correlations and the regression equations were computed by the Columbia Statistical Bureau.

TABLE V

MULTIPLE CORRELATIONS AND REGRESSION EQUATIONS OF SELECTED GROUPS OF TESTS WITH FACTORS

$$R_{V \bullet 1.4.6.10.12} = .7756$$
$$V = .3231\,X_1 + .2688\,X_4 + .1986\,X_6 + .1350\,X_{10} + .0799\,X_{12} - .2726$$

$$R_{N \bullet 2.5.8.11} = .6616$$
$$N = .1664\,X_2 + .2370\,X_5 + .4130\,X_8 - .0194\,X_{11} + 10.1499$$

$$R_{S \bullet 3.7.13.14} = .8278$$
$$S = .2258\,X_3 + .0721\,X_7 + .2676\,X_{13} + .4320\,X_{14} + .1265$$

TABLE VI

DISCRIMINATION OF CRITERION GROUPS BY FACTOR SCORES*

Criterion	Factor	Mean of Upper group	Mean of Lower group	D	σ_D†	$\frac{D}{\sigma_D}$
Total grades	V	55.20	45.97	9.23	2.56	3.61
	N	53.97	47.33	6.64	2.05	3.24
	S	52.70	48.33	4.37	3.04	1.44
Social science	V	52.77	46.23	6.54	3.12	2.10
	N	51.10	47.50	3.60	2.58	1.40
	S	49.17	47.43	1.74	3.23	.54
Science	V	56.97	44.57	12.40	2.37	5.23
	N	56.07	43.87	12.20	2.21	5.52
	S	56.00	40.90	15.10	2.59	5.83
Mathematics	V	52.63	45.57	7.06	2.80	2.52
	N	55.37	46.33	9.04	2.25	4.02
	S	53.97	43.63	10.34	2.84	3.64
Science & mathematics	V	54.17	45.47	8.70	2.52	3.46
	N	56.87	45.30	11.57	2.27	5.10
	S	54.13	42.00	12.13	3.11	3.90
Foreign language	V	53.60	46.97	6.63	2.70	2.46
	N	51.83	48.70	3.13	2.31	1.35
	S	48.33	49.73	—1.40	2.67	—.52
English	V	53.47	45.93	7.54	2.77	2.72
	N	50.90	49.87	1.03	2.39	.43
	S	50.20	47.90	2.30	3.01	.76
Foreign language & English	V	54.87	46.63	8.24	2.36	3.49
	N	52.20	48.50	3.70	2.55	1.45
	S	49.27	48.70	.57	2.99	.19

* Each criterion group consists of 30 students. The upper and lower criterion groups consist, then, respectively, of the best and the poorest sixths of the students in a given scholastic category.

† The use of Student's distribution would have given no essentially different results because of the size of the samples.

TABLE VIa

INSTANCES IN WHICH CRITERION GROUPS WERE SIGNIFICANTLY SEPARATED BY FACTOR SCORES

Discriminated by	Criterion	$\frac{D}{\sigma_D}$	
Verbal factor	Total grades	3.61	Significant*
	Social science	2.10	98 chances in 100
	Science	5.23	Significant
	Mathematics	2.52	99.4 chances in 100
	Sci. & math.	3.46	Significant
	Foreign lang.	2.46	99.3 chances in 100
	English	2.72	99.7 chances in 100
	For. lang. & Eng.	3.49	Significant
Numerical factor	Total grades	3.24	Significant
	Science	5.52	Significant
	Mathematics	4.02	Significant
	Sci. & math.	5.10	Significant
Spatial factor	Science	5.83	Significant
	Mathematics	3.64	Significant
	Sci. & math.	3.90	Significant

* i.e., critical ratio is greater than 3, the generally accepted criterion of a reliable difference.

superior and inferior students in each grade category are given in Table VI.

It will be seen that eleven of the twenty-four critical ratios are greater than 3, the generally accepted criterion of a statistically reliable difference. Fisher (4) has suggested that differences which have 95 chances in 100 of being real possess 'fiducial' reliability (i.e., are sufficiently great to be seriously considered as being reliable). If we accept this criterion we may add four more discriminations to our list of reliable differences: the chances are 98 in 100 that there is a true difference in verbal ability, i.e., the verbal factor, between superior and inferior students in the social sciences; 99.4 in 100 that there is a true verbal ability difference between

superior and inferior students in mathematics; 99.3 in 100 that there is a true verbal ability difference between superior and inferior students in foreign language, and 99.7 in 100 that there is a true verbal ability difference between superior and inferior students in English courses. In all, then, we may say that fifteen of the twenty-four differences are statistically significant. For convenience, we have appended Table VIa which lists the reliable discriminations by the verbal factor, the numerical factor and the spatial factor.

CHAPTER III

Discussion of Results

From the statistically significant critical ratios it may be concluded that there is a definite relationship between scholastic achievement and the factors we have dealt with in this study. Good and poor students in all of the scholastic categories are satisfactorily discriminated by at least one of the factors, and this despite the fact that our estimations of the factor scores are based on multiple correlations which are none too high, the highest being only .83. The probabilities are very high that the discriminations would have been even more satisfactory if we had been able to secure better estimations of abilities in the factors, since the chance element in any scores does not generally favor statistically significant differences. We may therefore conclude that the factors have psychological validity in the same sense that any test can be said to have psychological validity when evaluated against an external criterion.

Variations in the sizes of the critical ratios from one scholastic category to another may be a function of two factors: (1) the intrinsic relationship of the factor to the abilities involved in any particular line of scholastic achievement, and (2) the varying validities of school grades as measures of ability in the different lines of scholastic achievement.

It is because of this second factor that we cannot conclude from our table of discriminations that the verbal factor is involved to a

greater extent in the sciences than it is involved in the study of English or foreign languages. Indeed, granting the validity of the factor scores, we may utilize them to evaluate the relative validity of grades in the different scholastic categories as measures of ability (and, since there is no offhand reason to suppose that factors other than ability which determine school achievement are differentially involved in the different scholastic categories, as measures of performance). The argument may be stated in this way. If verbal factor scores discriminate between good and poor students in science, as they do, to the extent of the chances being 999, 999 in a million in favor of a true difference; and if, as is also the case, verbal factor scores are estimated solely from verbal tests; then, verbal ability must be a significant factor in achievement in science courses and such scores must be valid indices of verbal ability. If, now, these scores discriminate to a much lesser extent between good and poor students in language courses (where we would expect verbal ability to be at a premium), then common sense would tell us that it is the evaluation of the students in their courses that is at fault. The fact that the verbal factor discriminates between good and poor students selected on the basis of performance in both foreign language and English courses better than it does between good and poor students selected on the basis of English courses or foreign language courses alone tends to support this conclusion. Assuming that both types of courses involve verbal ability to approximately the same extent and that the primary ability determinant of grades in these courses (as we will indicate below) is verbal ability, then errors of evaluation in one category would tend to be balanced out by errors of evaluation in the other. A third category, then, based on these two, would tend to possess greater validity. The larger number of courses involved in the third category would, of course, also tend to make the evaluations more reliable and, hence, of greater theoretical validity.

On this basis, we may tentatively rank the different scholastic categories for the validity of grades as indices of ability as follows: science, mathematics, science and mathematics, combined, total grades, foreign language and English, combined; with English, foreign language and social science vying for last place.

On the other hand, we may within a single scholastic category evaluate the relative significance of the different factors for achievement in that category, since the degree of validity of the grades remains constant within the category. Unfortunately, the differential reliabilities of the estimates of the different types of factor scores might tend to obscure or invalidate this comparison. Yet such a comparison seems to make psychological sense in our data.

Bearing in mind, then, the possible source of error, we may tentatively state that considering all college courses in a single group, the verbal factor is the most important in differentiating good and poor groups of students, with the numerical factor a close second and the spatial factor playing at best a minor role. In the social sciences, the verbal factor is most important, the numerical factor plays at best a minor role, and the spatial factor none at all. In the science courses, all three factors are of about equal importance. In the mathematics courses, the numerical factor is the most important, followed in order by the spatial and the verbal. Considering science and mathematics together, the most important is the numerical factor, followed in order by the spatial and the verbal. In all three categories of language courses, the only important factor, of those we are considering, is the verbal.

SUMMARY

1. The purpose of this study is to find the relation of verbal, numerical, and spatial factors to scholastic achievement.

2. The results of a battery of 14 tests (5 verbal, 4 numerical, 4 spatial and an anagrams test) given to 186 students of the College of the City of New York by Dr. G. M. Smith were factorized by Thurstone's 'center of gravity' method and the resulting factors rotated to give the best possible statistical and psychological pattern. Factor scores were estimated with the aid of multiple regression equations for the verbal, numerical, and spatial factors.

3. Scholastic records of 178 of these students were obtained and their scholastic performance was analyzed into eight categories: total grades, social science, science, mathematics, science and mathe-

matics, combined, foreign language, English, and foreign language and English, combined.

4. The upper and lower sixth of the students in each of the scholastic categories were selected to serve as criterion groups. Factor scores were studied with reference to their ability to discriminate criterion groups.

5. Good and poor students in science, mathematics, and science and mathematics courses, combined were significantly separated by each of the factors; good and poor students selected on the basis of total grades were significantly separated by both the verbal and numerical factors; criterion groups in the four remaining scholastic categories were discriminated by the verbal factor.

6. The grades in the different scholastic categories were evaluated, on the basis of the sizes of the critical ratios, from the point of view of their validity as measures of ability. Thus, the violation of our expectations that the verbal factor should play a larger role in differentiating good and poor language students than in differentiating good and poor science students may be attributable either to the lack of validity of the factors, or, to the lower validity of language grades than of science grades. Since, however, the factor scores have already been shown to be valid, we must assume that the discrepancy is due to the fact that language grades are less valid than science grades.

7. An evaluation was made of the relative significance of each factor for achievement in each of the scholastic categories.

BIBLIOGRAPHY

1. Asher, E. J. The predictive value of mental tests that satisfy Spearman's tetrad criterion. *J. Appl. Psychol.*, 1929, *13*: 152-158.
2. Brigham, C. C. A study of error. *College Entrance Examination Board,* 1932. Chaps. II and VIII.
3. Brolyer, C. R. Fifth annual report of the Commission on Scholastic Aptitude. *College Entrance Examination Board,* 1930, 1-28.
4. Fisher, R. A. The design of experiments. Edinburgh, Oliver and Boyd, 1935.

5. Flanagan, J. C. Factor analysis in the study of personality. *Stanford University Press,* 1935. Pp. 103.
6. Garrett, H. E. Differentiable mental traits. *Psychol. Record,* 1938, *2*:259-298.
7. Guilford, J. P. *Psychometric methods.* McGraw Hill, 1936.
8. Guilford, J. P., and Guilford, R. B. An analysis of the factors in a typical test of introversion-extroversion. *J. Abnorm. Soc. Psychol.,* 1934, *28*:377-399.
9. Marshall, A. J. The standardization of Spearman's "measure of intelligence" for Perth, Western Australia. *Aust. Council Educ. Res. Ser.,* 1934. No. 22:23-50.
10. Schneck, M. M. R. The measurement of verbal and numerical abilities. *Arch. Psychol.,* 1929. No. 107. Pp. 49.
11. Smith, G. M. Group factors in mental tests similar in material or in structure. *Arch. Psychol.,* 1933. No. 156. Pp. 56.
12. Thomson, G. H. The factorial analysis of human ability. *Houghton Mifflin Company,* 1939. Pp. xv + 326.
13. Thurstone, L. L. The vectors of mind. *University of Chicago Press,* 1935. Pp. xv + 266.
14. Thurstone, L. L. Shifty and mathematical components. *Psychol. Bull.,* 1938, *35*:223-236.
15. Thurstone, L. L. Primary mental abilities. *Psychometric Society,* 1938. Pp. 121.
16. Walters, E. H., and Thomas, F. C. Some notes on the standardization of Professor Spearman's "measure of 'intelligence' for use in schools." *Forum Educ.,* 1929, *7*:35-42.

APPENDIX A

The Tests

For the convenience of the reader, sample items of the tests used are here cited from Smith (11).

The following tests are sufficiently familiar not to require illustration: (1) A sentence completion test, of the selective type;

(2) a number series completion test; (3) the Minnesota Paper Form Board Test which requires the subject to indicate with pencil lines how several small two-dimensional figures may be fitted into certain larger areas; (8) an arithmetic reasoning test; (9) an anagrams test; and (14) Kelley's spatial test, modified to make the items more difficult and the scoring more objective, in which the task is similar to that in the Minnesota Paper Form Board, but in which the ability to determine in imagination which of certain elementary shapes need to be turned over (not merely rotated in the same plane) in order to fit the given larger figures is also required.

The analogies tests require the subject to pick out the word, number or figure from the parentheses which stands in the same relation to the third word, number or figure as the second stands to the first. Below are sample items:

Verbal

No. 30. plumber pipes hammer (A carpenter B saw C nails D house) Answer (C)

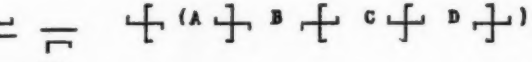
Numerical

No. 30. 3 1/27 2 (A ¼ B 0.125 C 1/16 D 0.25) Answer (B)

Grammatical

No. 31. I us he (A his B they C them D we) Answer (C)

Spatial

No. 34.  Answer (B)

The generalizations tests require the subject to select the three numbers, words or figures which have some specific quality in common or which have the most specific qualities in common. Below are sample items:

Verbal

No. 28. A youth B reporter C sapling D cub E bear
Answer (ACD)

Numerical

No. 25. A 9 B 27 C 45 D 81 E 24 Answer (ABD)

Grammatical

No. 31. A damn B sky C heavens D blue E oh
Answer (ACE)

Spatial

No. 30.

Answer (ACD)

For fuller details concerning the construction of these tests, the reader is referred to Smith (11).